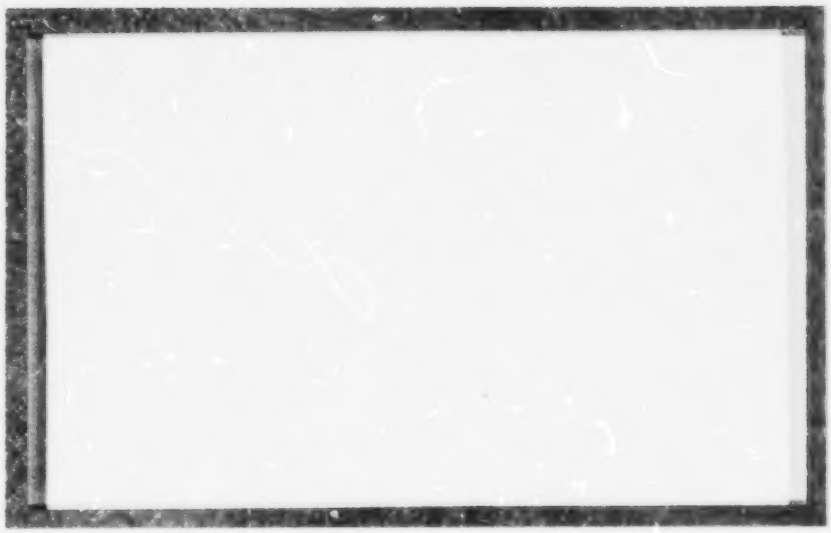
UNIVERSITÉ DE MONTRÉAL



CENTRE DE RECHERCHE SUR LES TRANSPORTS

CASE POSTALE 6128, SUCCURSALE A, MONTRÉAL, QUÉBEC, H3C 3J7

A ROUTING ALGORITHM FOR POSTAL TRUCKS IN URBAN AREAS*

by

Gilbert Laporte[1]
Suzanne Chapleau[1]
Philippe-Eric Landry[2]
and
Hélène Mercure[3]

1 Centre de recherche sur les transports, Université de Montréal, C.P. 6128, Succursale A, Montréal H3C 3J7, Canada.

2 Société canadienne des postes, 545 rue MacArthur, Ville Saint-Laurent H4T 1T6, Canada.

3 École des Hautes Études Commerciales de Montréal, 5255 avenue Decelles, Montréal H3T 1V6, Canada.

* The authors are grateful to the Canadian Natural Sciences and Engineering Research Council (grant A4747) and to the Canada Post Corporation for their support. Thanks are also due to the anonymous referees for their valuable comments.

Cette étude a été publiée grâce à une subvention du fonds F.C.A.R. pour l'aide et le soutien à la recherche.

Université de Montréal
Centre de recherche sur les transports - Publication #533
July 1987

Dépôt légal - 3e trimestre 1987

ABSTRACT

Mailbox collection constitutes a complex and costly operation in most postal services. The authors recently undertook a study aimed at improving this operation at the Canada Post Corporation. This paper describes an improvement to the clustering and routing algorithm currently in use. It is based on the exact solution of a generalized version of the travelling salesman problem in the last phase of the routing problem. The results obtained so far indicate that significant savings can be achieved.

Key words : mailbox collection, vehicle routing, travelling salesman problem.

RÉSUMÉ

La levée des boîtes à lettre constitue une opération complexe et coûteuse dans la pluspart des services postaux. Les auteurs ont entrepris une étude visant à améliorer cette opération à la Société canadienne des postes. Cet article décrit une amélioration à l'algorithme de regroupement et de construction de tournées présentement en usage. Cet algorithme repose sur la résolution exacte d'une version généralisée du problème du voyageur de commerce, dans la dernière phase de l'algorithme. Les résultats obtenus jusqu'ici indiquent qu'on peut réaliser des économies appréciables.

Mots-clés : levée des boîtes à lettre, problèmes de tournées, problème du voyageur de commerce.

1. INTRODUCTION

Mailbox collection constitutes a complex and costly operation in most postal services. The authors recently undertook a study aimed at improving this operation at the Canada Post Corporation (CPC). The results obtained so far indicate that significant savings can be achieved.

Routing problems arising from postal services operations can be classified into two broad categories: **arc routing** problems (e.g. door-to-door mail delivery — for a recent survey, see Levy (1987)) and **node routing** problems, such as the mailbox collection problem treated in this paper. For a survey of node routing methodology, see Bodin et al. (1983) and Laporte and Nobert (1987).

Designing optimal mailbox collection routes usually involves two main steps:

(i) **Clustering.** Determine which mailboxes will belong to each vehicle route. Typically, each cluster will include between 20 and 50 mailboxes. One important consideration in the design of clusters is that they should be relatively compact; as a side effect, the collection routes generated within the clusters will tend to be shorter.

(ii) **Routing.** Generate a minimum length vehicle route through the mailboxes of each cluster.

Traditionally, these two phases have been executed sequentially and somewhat independently by CPC, leading to an overall satisfactory but sub-optimal solution. Given the size of the problem, we feel this is a sensible approach, although some local improvements are possible. This paper describes such an improvement. It was developed by the authors over the last two years and has led to travel distance reductions of up to 14.5% on 10 test problems.

The remainder of the paper is organized as follows. In section 2 and 3, we define the distance matrix and we summarize the clustering and routing algorithm used by CPC. Section 4 gives the mathematical programming formulation of the routing problem. Section 5 summarizes the proposed routing algorithm and finally, we present in section 6 some computational results and comparisons. The conclusion follows.

2. DEFINING THE DISTANCE MATRIX IN A POSTAL CONTEXT

The first step of most routing algorithms usually consists of identifying all stops and of computing the shortest distance matrix between these stops. While this task is relatively straightforward in most applications, it can be rather tricky in a postal context where (i) there are one-way streets, U-turn interdictions and some left-turn prohibitions; (ii) some mailboxes are located along the edge of a street while others are located at street intersections; (iii) for safety reasons, postal truck drivers do not cross a street on foot in order to empty a mailbox. This last constraint creates some difficulty in the definition of the cost matrix since the shortest distance between two mailboxes is no longer uniquely defined, but depends rather on the itinerary followed by the postal truck before arriving at the first mailbox.

Consider, for example, the street plan shown in figure 1. Assume there are no one-way streets and no left-turn interdictions. Consider two mailboxes located at sites A and B. Mailbox A can be served either by a truck arriving from the South, or by a truck arriving from the West. In the first case, the shortest distance $c'_{A,B}$ from A to B is simply the length of the vertical arrow from A to B. However, in the second case, the truck will have to go round the block before reaching B. Then, the shortest distance from A to B, $c''_{A,B}$. A way round to this problem is to duplicate

all mailboxes located at intersections. Thus, mailbox A would be replaced by A1, located on arc AD and by A2, located on arc CA. This yields a partition of all stops into three sets:

I all stops not located at street corners;

J_1 all first copies of stops originally located at intersections:

J_2 all second copies of stops originally located at intersections.

The problem can now be slightly simplified as follows. Let an arc be defined as one side of a street between two intersections. Replace every arc by a node and define the distance from arc (i,j) to arc (k,l) as the shortest distance $c_{j,k}$ from j to k.

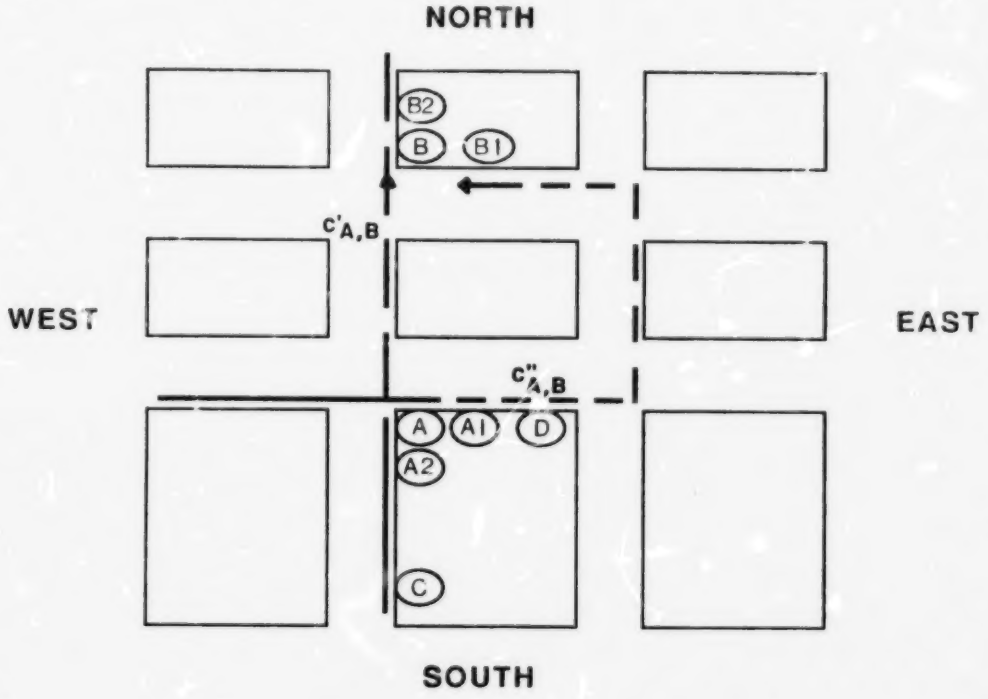


Figure 1: Street plan showing duplicated mailboxes

3. THE CLUSTERING AND ROUTING ALGORITHM

We summarize in this section the clustering and routing procedure in use at CPC.

Step 1 **Definition of a distance matrix.** Duplicate mailboxes located at street intersections and define the distance matrix, as explained in section 2.

Step 2 **Initial clustering.** This step consists of defining clusters corresponding to all stops included on a single vehicle route. Cluster C_p will include n_p stops. Clusters are gradually built around the mailbox which is furthest from the post office and which is not already included in a cluster. A cluster is completed when it contains enough mailboxes or when the approximate time required to service these mailboxes is reached. At the end of this step, all clusters have been defined, but they may not be compact.

Step 3 **Updated clustering.** Determine the centroid of each cluster, i.e. the mailbox for which the sum f_p of squared distances to all remaining mailboxes in the cluster is minimized. As in Bourjolly et al. (1981), perform marginal modifications to the clusters (and while doing so, redefine the centroids) as long as $\sum_p f_p$ can be reduced. These modifications are of two kinds: (i) simple moves of stops from one cluster to an adjacent one; (ii) exchanges between two adjacent clusters. At the end of this step, we have obtained compact clusters with respect to the distance matrix defined in step 1, although they may not be visually compact.

Step 4 **Route generation.** Within each cluster, use a nearest neighbour heuristic to generate a route going through all mailboxes originally located on street edges and one copy of every box located at an intersection. The two ends of the route usually correspond to two

points located near the "extremities" of the cluster. They are often specified by the user. Improve this route by means of the 2-opt procedure used in Bodin and Berman (1979).

Step 5 **Updated clustering.** Check whether the total distance travelled can be reduced by performing marginal modifications to the clusters, as in step 3. This procedure is repeated as long as distance reductions can be achieved. This clustering method tends to produce solutions having a small inter-stop distance, but may involve a large amount of line-haul travel and does not always minimize total distance. See Daganzo (1984) and Langevin and Soumis (1987) for theoretical results on the question of cluster shapes.

Step 6 **Route reoptimization.** In this step, duplicated mailboxes are replaced by their copy if this saves distance. Formally, this is done as follows. Initially, each route can be described by a sequence of n'_p stops of the form $(i_1,\ldots,i_{n'_p})$. Now consider all subsequences of the form $(i_s, i_{s+1},\ldots,i_t)$ where i_s, $i_t \in I$, $i_{s+1},\ldots,i_{t-1} \in J_1 \cup J_2$. Let i'_j denote the copy of i_j (if $i_j \in J_1$, then $i'_j \in J_2$, and conversely). A complete search is made to determine whether distance reductions could be achieved by replacing any i_j by i'_j for $s+1 \leq j \leq t-1$. Since all possibilities are examined, this involves a total of 2^{t-s-1} evaluations. Fortunately, in practice, the value of t-s-1 is never very large (it rarely exceeds 5).

4. GENERALIZED TRAVELLING SALESMAN PROBLEM FORMULATION

One disadvantage of the above procedure is that once routes have been generated after step 5, they can only be improved marginally by

replacing duplicated mailboxes by their copy. The relative order of stops stays however the same. CPC felt that an improvement could be achieved if it was possible to integrate the route generation procedure and the selection of the "best copy" of all mailboxes originally located at street intersections. We show that, provided the cluster cardinalitites are kept relatively small, this problem can be solved to optimality with reasonable computing effort.

Consider the original problem and define

(1) for each arc a (corresponding to one side of a street) where one or more mailboxes are located, a set $S_a = \{a\}$;

(2) for each mailbox ℓ located at an intersection, a set $T_\ell = \{\ell, \ell'\}$ where ℓ and ℓ' are the two copies of the mailbox.

Also define $S = \bigcup_a S_a$ and $T = \bigcup_\ell T_\ell$. At the end of this step, we have created n' sets and $r = |S| + |T|$ nodes. Our aim is to determine the shortest route starting and ending at two given nodes, and passing through one node of every set. Introduce a dummy node i_0 with the corresponding set $V = \{i_0\}$. Define a zero distance between i_0 and the two end nodes and infinite distance between i_0 and all other nodes. The problem is then a **generalized travelling salesman problem** (GTSP) through $n = n' + 1$ sets of nodes.

The GTSP constitutes an extension of the travelling salesman problem (TSP) (see Dantzig et al. (1954) for example); in the TSP, all sets of nodes have a cardinality of one. The GTSP has been studied by Laporte and Nobert (1983) for symmetrical distance matrices and more recently by Laporte et al. (1987) for asymmetrical distance matrices, as is the case in this application.

The formulation of the (asymmetrical) GTSP particularized to our problem is as follows. Let $N = S \cup T \cup V$ be the set of all nodes and $C = (c_{i,j})$ the matrix of shortest distance between pairs of nodes of N. Let $x_{i,j}$ be 0-1 variables indicating whether arc (i,j) belongs to the optimal solution ($x_{i,i} = 0$) or not ($x_{i,i} = 1$). Variables $x_{i,i}$ are only defined for $i \in T$ and must be interpreted as zero otherwise. The problem is then to

$$\text{minimize} \quad z = \sum_{\substack{i,j \in N \\ i \neq j}} c_{i,j} \, x_{i,j} \qquad (P)$$

subject to

$$\sum_{i \in N} x_{i,k} = \sum_{j \in N} x_{k,j} = 1 \qquad (k \in N) \qquad (2)$$

$$\sum_{i \in T_\ell} x_{i,i} = 1 \qquad (\text{all } T_\ell) \qquad (3)$$

$$\sum_{\substack{i,j \in M \\ i \neq j}} x_{i,j} \leq |M| - 1 \qquad (M \subset N;\ M \neq T_\ell,\ \text{all } \ell) \qquad (4)$$

$$x_{i,j} = 0 \text{ or } 1 \qquad (i,j \in N) \qquad (5)$$

In this formulation, constraints (2) ensure that every mode of $S \cup V$ is entered and left once and exactly once. When applied to nodes k of T, these constraints ensure that k is entered and left once if and only if $x_{k,k} = 0$, i.e. if k appears in the optimal solution. Constraints (3) specify that in every T_ℓ, one node out of two is visited. Constraints (4) are the classical subtour elimination constraints originally developed by Dantzig et al. (1954) for the TSP. They specify that in any subset of nodes of N, the number of arcs must be strictly less than the number of nodes, thus preventing the formation of tours involving less than $|N|$

nodes. It is not necessary to impose such constraints for $M = T_\ell$ since the two nodes of T_ℓ can never enter the solution simultaneously. Finally, constraints (5) specify the binary conditions on the variables.

5. ALGORITHM

The algorithm used for the solution of (P) was derived from the GTSP algorithm described in Laporte et al. (1987). This algorithm (step 7) is executed after step 6 in the procedure outlined in section 3. It is applied to each cluster C_p independently and is used as a post-optimizer, once clusters guaranteeing reasonable route lengths have been defined. However, since the GTSP algorithm can only solve relatively small size problems in a reasonable time, we have opted to apply it to clusters involving up to about 30 nodes, corresponding roughly to 20 original mailbox locations, 10 of which are duplicated. With this in mind, two approaches are possible for the definition of the clusters in step 2 of the algorithm: (i) ensure that $|C_p| \leq 20$ for all p and later generate complete vehicle routes by merging two or three of the shorter routes; (ii) define the clusters in such a way that they include as many mailboxes as normally constitute a vehicle route; since in most cases these routes will involve around 45 different stops, they will have to be broken down into two or three segments before applying the GTSP algorithm. We have opted for the second approach after having observed that (long) vehicle routes very often have "natural breaking points", corresponding to relatively long driving stretches between small clusters of stops.

The GTSP algorithm was constructed by extending and otherwise modifying the Carpaneto and Toth (1980) branch and bound algorithm for asymmetrical TSP's. At every node of the search tree, this algorithm solves an assignment problem (AP) in which some variables are forced to take the value 0 or 1. At a given node h, a check is made to determine whether the current solution is legal or not. A solution may be illegal for one or

both of the following reasons: (i) at least one T_ℓ is not visited; (ii) illegal subtours exist. Illegal solutions are eliminated by generating subproblems in which the appropriate variables are fixed at 0 or at 1 (we used here the Carpaneto and Toth partitioning rule). The remainder of the algorithm follows the usual branch and bound methodology.

6. COMPUTATIONAL RESULTS

The algorithm was coded in FORTRAN and tested over 10 clusters in two postal regions of the Montreal area. Computational results are reported in table 1. The results indicate that the GTSP algorithm can be applied provided the number of nodes in a cluster is not too large. Memory capacity ($250{,}000_8$ words) was exceeded in two problems (2 and 7) involving 27 and 26 nodes respectively. These are the two problems where the largest number of mailboxes were duplicated; in such problems, the number of time constraints (3) were violated in the course of the branch and bound algorithm tended to be high. This generated an excessive number of subproblems and the algorithm could not be completed. Otherwise, the results are very positive. Computation times are in most cases reasonably low and the distance savings due to the introduction of the GTSP algorithm is in some cases relatively high; since mailbox collections are carried out nationwide several times a day, this can represent substantial savings in the long run. The discrepancies appearing in column 7 of table 1 (percentage improvement) were examined, but could not be related in any obvious way to some problem characteristics. They constitute more a reflection of the degree of suboptimality of the CPC algorithm than of the quality of the GTSP algorithm which performs consistently well and always provides the optimum, when run to completion.

Table 1: Summary of Computational Results

Cluster	Number of mailboxes	Number of nodes after duplications	Solution value (in meters) after step			Percentage improvement [(6)-(7)]÷(6)	CPU time (seconds) (*)
			5	6	7		
1	14	22	15923	14397	13931	3.2%	94
2	15	27	15713	15449	15449	0.0%	(**)
3	13	22	14293	14293	12225	14.5%	252
4	10	18	9219	7942	7671	3.4%	31
5	19	26	8643	8643	8643	0.0%	18
6	12	20	7735	7290	7290	0.0%	19
7	15	26	12204	11642	11237	3.4%	(**)
8	11	17	10297	10140	9644	4.9%	23
9	13	22	9837	9285	8000	13.8%	31
10	13	24	11948	10870	9808	9.7%	80

(*) Université de Montréal CYBER 173 computer.
(**) Memory requirements exceeded. Optimality not proven.

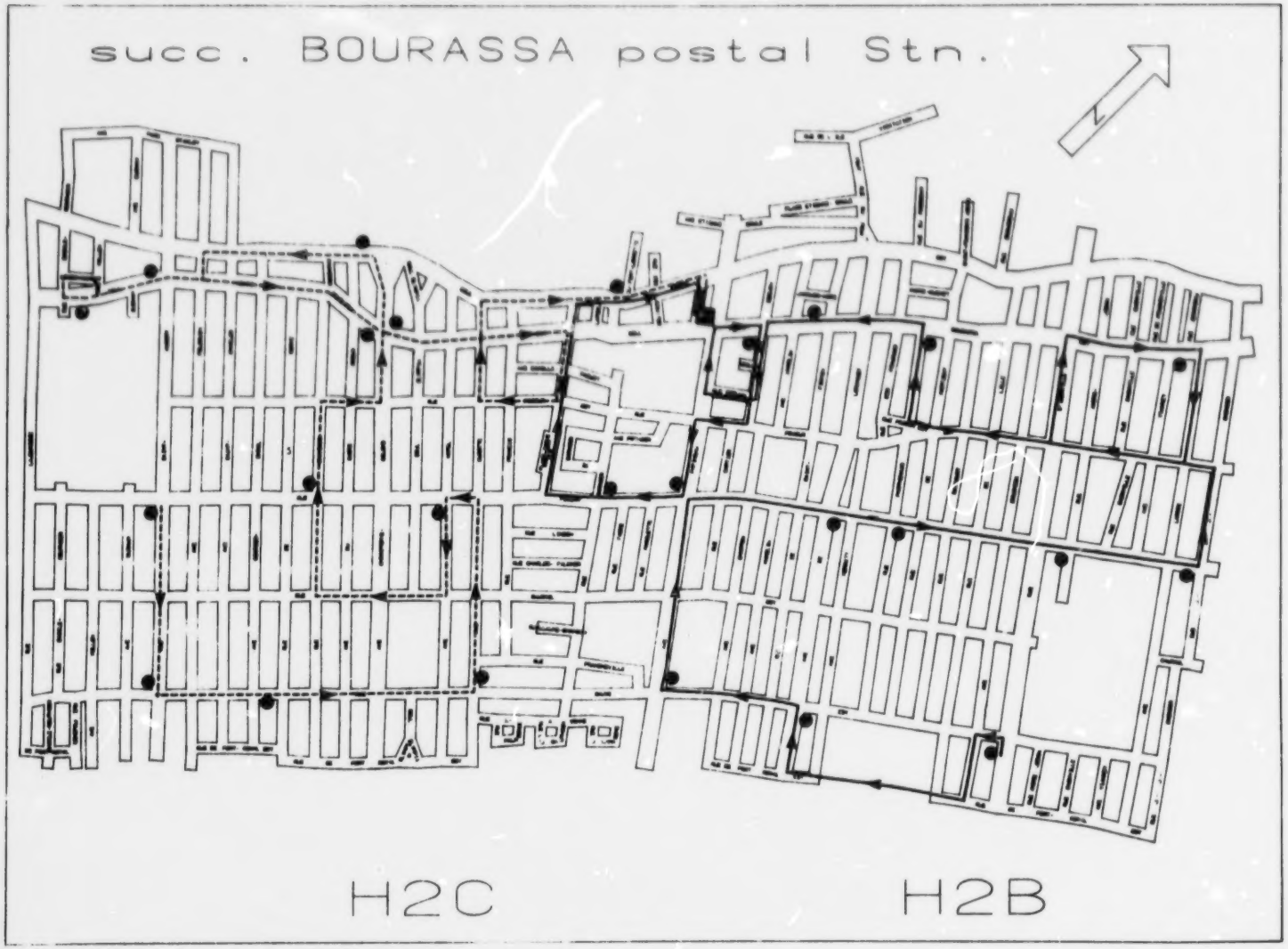
Finally, some of the itineraries produced by the algorithm were examined. An example is provided in figure 2. This diagram shows that the collection route can become at times quite intricate. This is due to the presence of the large number of one-way streets and left-turn interdictions in a city such as Montreal. The whole procedure described in this paper was developed by assuming fixed mailbox locations. By examining the two routes illustrated in figure 2, one can see several instances where minor changes in most mailbox locations (across the street, for example) would yield large distance savings. It is relatively straightforward to include in the GTSP algorithm the possibility of determining optimal mailbox locations (on either side of the street, for example, see Labbé and Laporte (1987)), together with the associated optimal routing. It seems, however that more often than not, mailbox relocations meet with some opposition from businesses and citizen groups concerned and that, for the time being, this issue should better be left untouched.

7. CONCLUSION

We have described in this paper a revised algorithm for the mailbox collection problem. No actual implementation has yet taken place but limited computational tests indicate that its use could result in substantial savings at CPC. The algorithm is also well-suited to integrate mailbox location to route design. A further related problem now investigated by our team is the optimal location of relay-(or transfer) boxes used by letter-carriers. Since mailboxes and relay-boxes are usually visited in one trip by the same truck, the two location-routing problems should be solved jointly.

FIGURE 2

Itinerary determined by the improved algorithm in clusters 9 (dotted line) and 10 (full line)



REFERENCES

BODIN, L., BERMAN, L. (1979) "Routing and Scheduling of School Buses by Computer", Transportation Science 13, 113-129.

BODIN, L.D., GOLDEN, B.L., ASSAD, A., and BALL, M. (1983) "Routing and Scheduling of Vehicles and Crews: The State of the Art", Computers and Operations Research 10, 69-211.

BOURJOLLY, J.M., LAPORTE, G., ROUSSEAU, J.-M. (1981) "Découpage électoral automatisé : application à l'Ile de Montréal", INFOR 19, 113-124.

CARPANETO, G., TOTH, P. (1980) "Some New Branching and Bounding Criteria for the Asymmetrical Travelling Salesman Problem", Management Science 26, 736-743.

DAGANZO, C.P. (1984) "The Distance Traveled to Visit N Points with a Maximum of C Stops per Vehicle: An Analytical Model and an Application", Transportation Science 18, 331-350.

DANTZIG, G.B., FULKERSON, D.R., JOHNSON, S.M. (1954) "Solution of a Large Scale Travelling Salesman Problem", Operations Research 2, 393-410.

LABBÉ, M., LAPORTE, G. (1986) "Maximizing User Convenience and Postal Service Efficiency in Post Box Location", Belgian Journal of Operations Research, Statistics and Computer Science 26, 21-36.

LANGEVIN, A., SOUMIS, F. (1987) "Design of Multiple-Vehicles Delivery Tours Satisfying Duration Constraints", Publication #513A, Centre de recherche sur les transports, Université de Montréal.

LAPORTE, G., NOBERT, Y. (1987) "Exact Algorithms for the Vehicle Routing Problem", in S. Martello, G. Laporte, M. Minoux and C. Riberio editors, Surveys in Combinatorial Optimization, Annals of Discrete Mathematics 31, North-Holland, Amsterdam, 147-184.

LAPORTE, G., MERCURE, H., NOBERT, Y. (1987) "Generalized Travelling Salesman through n Sets of Nodes: The Asymmetrical Case", forthcoming in Discrete Applied Mathematics.

LAPORTE, G., NOBERT, Y. (1983) "Generalized Travelling Salesman through n Sets of Nodes: An Integer Programming Approach", INFOR 21, 61-75.

LEVY, L.S. (1987) The Walking Line of Travel Problem: An Application of Arc Routing and Partitioning, Ph.D. Dissertation, College of Business and Management, University of Maryland at College Park, MD.